JA 606

Genet.; 49(3): 375–383 (1989)

ON ESTIMATION OF HARVEST INDICES

MURARI SINGH*

International Crops Research Institute for the Semi-Arid Tropics, Patancheru Andhra Pradesh 502324

(Received: March 5, 1988; accepted: January 9, 1989)

ABSTRACT

he conventional method of analysis of harvest indices consists of obtaining mean ratios o lot basis and applying analysis of variance to these ratios. These ratios need not confor e assumptions of the analysis of variance while the variables in the numerator an enominator may be in agreement with them. The application of analysis of variance of ratio a plot basis is criticized and an alternative procedure of estimating harvest index using rat f means is proposed. The comparison of the proposed estimate of harvest index with th nventional estimate has been made and is illustrated with field data from experiment nducted on groundnut. The proposed estimate (based on ratio of means) is more preci nd has smaller bias compared to the conventional estimate of harvest index.

ey words: Analysis of variance, ratio of random variables, harvest index.

There are numerous situations in agricultural research, especially in plan ng and plant physiology, where the experimenter is interested in assessing th e performance of a set of genotypes by two character means and their ratio stance, an experimenter is interested in estimating the economic yield, tota ical yield, and harvest index of a genotype. For example, in groundnut, th mic yield is represented by pods and the ratio of pod yield to the total yiel ass) is the harvest index, while in sorghum, the grain weigh res the economic yield. Another example is the estimation of shelling percentag undnut.

The problem of estimation of several ratios as characters within an individua plot has been discussed in [1, 2] using the almost unbiased ratio estimator [3]. Th method [1, 2] uses a linear function of the ratio of the means and the mean of the ratios of the two character values measured on the plants randomly selected from the plot. These estimates from individual plots were subjected to analysis. In the present paper, we consider the estimation of ratio characters (e.g. harvest index) from a designed experiment where there is one more factor (such as, block in randomized block design) in addition to the factor of interest (such as, genotype or variety). The measurements of the two characters used in the estimation of ratios are assumed to be available on a few [illegible] of the plants taken randomly in

mean values which may vary differently [9]. The distribution of r_{ij} will not be normal. Hence the three assumptions, (i) additivity of factor effects, (ii) constancy of error variances, and (iii) normality of errors, are not tenable. Therefore, ANOVA should not be applied to models (1), (2) and (4) simultaneously. Further the use of model (4) to estimate harvest index of i-th cultivar $H_i = \zeta_{ix}/\zeta_{iy}$ leads to the biased estimate

$$\bar{r}_i = (1/s) \sum_j \bar{r}_{ij}$$

and the bias of $\bar{r}_i$ (given in the following section) also depends on block effects. Hereafter $\bar{r}_i$ will be referred to as the conventional estimate used by several authors mentioned above.

In this paper, we suggest an alternative estimate of harvest index H_i and compare it with the conventional estimate r_i for their biases and mean square errors. An approximate test for equality of harvest index is also given.

MATERIALS AND METHODS

ESTIMATION OF HARVEST INDEX AND APPROXIMATE STANDARD ERROR

In models (1) and (2), errors ξ_{ijx} and ξ_{ijy} have means 0 and variances σ^2_x and σ^2_y, respectively, and correlation coefficient ρ. ξ_{ijz} and $\xi_{i'j'z}$ are independent for $i \neq i'$ or $j \neq j'$ $(z=x,y)$. The joint distribution of the pair (ξ_{ijx}, ξ_{ijy}) can be assumed to be bivariate normal.

The estimates of ζ_{ix} and ζ_{iy} can be taken as least square estimates

$$\hat{\zeta}_{ix} = \sum_j x_{ij}/s; \quad \hat{\zeta}_{iy} = \sum_j y_{ij}/s$$

and those of σ^2_x, σ^2_y as residual mean squares from analysis of variances on the data of x and y as

$$\hat{\sigma}^2_x = \Sigma \Sigma \hat{\xi}^2_{ijx}/n_e$$

$$\hat{\sigma}^2_y = \Sigma \Sigma \hat{\xi}^2_{ijy}/n_e$$

where $n_e = (s-1)(v-1)$ error degree of freedom and $\hat{\xi}_{ijx}$, $\hat{\xi}_{ijy}$ least squares residuals

$$\hat{\xi}_{ijx} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$$

a plot. The genotypes are assessed for economic and total biological yield generally using their means and standard errors obtained from analysis of variance (ANOVA) on these characters individually. Tests of significance of contrasts of interest are carried out when these characters satisfy a number of assumptions [4, 5]: (i) the treatment and environmental effects (such as, block effects in randomised complete block design) must be additive and without interaction when evaluated on plot basis; (ii) the experimental errors must have a common variance; (iii) the experimental errors should be normally distributed; and (iv) the experimental errors must all be independent.

The practice followed by several workers [6–8] to estimate harvest index is to generate vector R of ratios r_{ij} $(= x_{ij}/y_{ij})$ of the two variates x_{ij}, y_{ij}, and obtain the mean of r_{ij} values over replicates for each genotype and its standard error from the analysis of variance applied to r_{ij}. Here x_{ij} and y_{ij} are, respectively, the economic yield (X) and total yield (Y) of the plot of the i-th genotype in j-th block. Let the design used be a randomised complete block design (RBD) with v genotype and s block ($i = 1, \ldots, v$; $j = 1, \ldots, s$).

In the light of analysis of variance applied to variables X and Y (when the underlying assumptions of ANOVA are satisfied), it is not recommended to apply analysis of variance to their ratio R(=X/Y), since then the assumptions of the ANOVA are not satisfied. This can be explained as follows. Consider the model

$$x_{ij} = \zeta_{ix} + \beta_{jx} + \xi_{ijx} \tag{1}$$

$$y_{ij} = \zeta_{iy} + \beta_{jy} + \xi_{ijy} \tag{2}$$

where ζ_{iz}, β_{jz} represent the performance of i-th genotype and the effect of j-th block for character z(z=x,y) and

$$\sum_j \beta_{jx} = 0, \quad \sum_j \beta_{jy} = 0$$

The models (1) and (2) are additive in the effects of genotypes (treatment factor) and blocks (environment factor). The errors ξ_{ijx}, ξ_{ijy} satisfy the assumptions (ii)–(iv). It is easy to see that the ratio

$$r_{ij} = x_{ij}/y_{ij} = (\zeta_{ix} + \beta_{jx} + \xi_{ijx})/(\zeta_{iy} + \beta_{jy} + \xi_{ijy}) \tag{3}$$

in general, cannot be written as an additive model,

$$r_{ij} = R_i + B_j + e_{ij} \tag{4}$$

$$(i = 1, \ldots, v; j = 1, \ldots, s).$$

The variance of r_{ij} [e_{ij} in (4)] (expression given in the following section) is not constant for all i, j, as the variance of ratio of random variables depends on their

mean values which may vary differently [9]. The distribution of r_{ij} will not be normal. Hence the three assumptions, (i) additivity of factor effects, (ii) constancy of error variances, and (iii) normality of errors, are not tenable. Therefore, ANOVA should not be applied to models (1), (2) and (4) simultaneously. Further the use of model (4) to estimate harvest index of i-th cultivar $H_i = \zeta_{ix}/\zeta_{iy}$ leads to the biased estimate

$$\bar{r}_i = (1/s) \sum_j \bar{r}_{ij}$$

and the bias of $\bar{r}_i$ (given in the following section) also depends on block effects. Hereafter $\bar{r}_i$ will be referred to as the conventional estimate used by several authors mentioned above.

In this paper, we suggest an alternative estimate of harvest index H_i and compare it with the conventional estimate r_i for their biases and mean square errors. An approximate test for equality of harvest index is also given.

MATERIALS AND METHODS

ESTIMATION OF HARVEST INDEX AND APPROXIMATE STANDARD ERROR

In models (1) and (2), errors ξ_{ijk} and ξ_{ijy} have means 0 and variances σ^2_x and σ^2_y, respectively, and correlation coefficient ρ. ξ_{ijx} and $\xi_{i'j'z}$ are independent for $i \neq i'$ or $j \neq j'$ $(z=x,y)$. The joint distribution of the pair (ξ_{ijx}, ξ_{ijy}) can be assumed to be bivariate normal.

The estimates of ζ_{ix} and ζ_{iy} can be taken as least square estimates

$$\hat{\zeta}_{ix} = \sum_j x_{ij}/s; \quad \hat{\zeta}_{iy} = \sum_j y_{ij}/s$$

and those of σ^2_x, σ^2_y as residual mean squares from analysis of variances on the data of x and y as

$$\hat{\sigma}^2_x = \Sigma \Sigma \hat{\xi}^2_{ijx}/n_e$$

$$\hat{\sigma}^2_y = \Sigma \Sigma \hat{\xi}^2_{ijy}/n_e$$

where $n_e = (s-1)(v-1)$ error degree of freedom and $\hat{\xi}_{ijx}$, $\hat{\xi}_{ijy}$ least squares residuals

$$\hat{\xi}_{ijx} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..$$

$$\hat{\xi}_{ijy} = y_{ij} - \bar{y}_{i} - \bar{y}_{.j} + \bar{y}..$$

where bar (–) denotes mean over the dot (.) position(s).

The correlation coefficient is estimated by

$$\hat{\rho} = (\Sigma \Sigma \hat{\xi}_{ijx} \hat{\xi}_{ijy}/n_e)/(\hat{\sigma}^2_x \hat{\sigma}^2_y)^{1/2}$$

We propose to estimate the harvest index H_i by the formula

$$\hat{H}_i = h_i = \hat{\zeta}_{ix}/\hat{\zeta}_{iy}$$

It may be noted that the estimate h_i of H_i can also be derived from the functional relation

$$\zeta_{ix} = H_i\zeta_{iy} \quad (i = 1, 2, \ldots, v)$$

Where ζ_{ix} and ζ_{iy} are expressed in model equations (1) and (2). The estimate h_i is based on the ratio of two normal variates. The confidence limits for h_i can be seen from the Fieller's theorem [10]. The estimator h_i is biased for estimating H_i.

The expressions for the biases $B(h_i)$, $B(\bar{r}_i)$ and mean square errors $M(h_i)$, $M(\bar{r}_i)$ of the two estimators, h_i and $\bar{r}_i$ of H_i, will be obtained following some results for infinite populations simplified for normal distribution [9].

If w_1 and w_2 are random variables with means μ_1, μ_2, variances σ^2_1, σ^2_2, and correlation ρ, then approximations to the expected value E(.) and mean square error M(.) of w_1/w_2 are

$$E(w_1/w_2) = (\mu_1/\mu_2)(1 + (C^2_2 - \rho C_1C_2)(1 + 3C^2_2)$$

$$M(w_1/w_2) = (\mu_1/\mu_2)^2(C^2_1 + C^2_2 - 2\rho C_1C_2 + 3)((1 + 2\rho^2)C^2_1C^2_2 + 3C^4_2 - 6\rho C_1C^3_2))$$

where $C_1 = \sigma_1/\mu_1$, $C_2 = \sigma_2/\mu_2$ are coefficients of variation of w_1 and w_2, respectively.

Applying the above results for h_i, we get after simplification

$$B(h_i) = H_i\alpha_1(1/s + 3C^2_y/s^2)$$

$$M(h_i) = H^2_i(\alpha_2/s + \alpha_3/s^2)$$

where

$$\alpha_1 = C^2_y - \rho C_xC_y, \quad \alpha_2 = C^2_x + C^2_y - 2\rho C_xC_y$$

$$\alpha_3 = 3(1 + 2\rho^2)C^2_xC^2_y - 18\rho C_xC^3_y + 9C^4_y$$

$$C_x = \sigma_x/\zeta_{ix}, \quad C_y = \sigma_y/\zeta_{iy} \quad (i = 1, 2, \ldots, v)$$

These expressions are up to the order of $(1/s^2)$. However, in many cases simplification up to the order of $(1/s)$ is adequate. In the example given in the next section, we find that the biases are negligible and there is no marked difference between the expressions evaluated up to the orders of $(1/s)$ and $(1/s^2)$. The estimates of bias and mean square error are obtained by substituting the estimates of parameters

We now evaluate bias and mean squared error for the conventional estimator r_i. Writing

$$\bar{r}_i = (1/s)\sum_j (x_{ij}/y_{ij}),$$

we have $E(\bar{r}_i) = (1/s)\sum_j E(x_{ij}/y_{ij}) = (1/s)\sum_j E(r_{ij})(1 + \alpha_1(1+3C^2_y))$.

Using the expansion of the binomial term with negative power in the expression of $E(r_{ij})$, we get

$$E(r_{ij}) = H_i(1 + \delta_{ijx} - \delta_{ijy} - \delta_{ijx}\delta_{ijy} + \delta^2_{ijy} + \ldots .)$$

where $\delta_{ijx} = \beta_{jx}/\zeta_{ix}$, $\delta_{ijy} = \beta_{jy}/\zeta_{iy}$.

Ignoring the terms in δ's with power higher than two, we have approximately

$$B(\bar{r}_i) = E(\bar{r}_i) - H_i$$

$$= H_i(\alpha_1(1 + 3C^2_y) + 1 + \alpha_2(1 + 3C^2_y))(\Sigma\delta_{ijy}{}^2/s - \sum_j \delta_{ijx}\delta_{ijy}/s)$$

The mean squared error is given by

$$M(\bar{r}_i) = (1/s^2)\sum_j M(x_{ij}/y_{ij})$$

$$= (1/s^2)\sum_j E(r_{ij})(\alpha_2 + \alpha_3)$$

$$= H^2_i(\alpha_2/s + \alpha_3/s)(1 + \sum_j(\delta^2_{ijx} + 3\delta^2_{ijy} - 4\delta_{ijx}\delta_{ijy})/s).$$

The biases and mean squared errors of h_i and $\bar{r}_i$ can be easily compared using the above expressions. Considering the leading term free from block effects in δ's, the bias and mean squared error for h_i are lower than that of $\bar{r}_i$ and decreases with increase in replications. Thus h_i is, therefore, better than $\bar{r}_i$ for estimating H_i. Finally, we compare the performance of h_i and $\bar{r}_i$ using data from an experiment conducted at ICRISAT.

COMPARING CULTIVARS FOR THEIR HARVEST INDICES

When the bias of the estimate of harvest index h_i is negligible its mean squared error equals its variance. For large residual degree of freedom, the variance estimate closely approaches the true variance. The estimate h_i is a consistent estimate of H_i (as can be seen from its mean square error expression). Thus following Rao [8], the statistic

$$Q = \sum_{i=1}^{v} (h_i - \bar{H})^2/M(h_i)$$

where $\bar{H} = \sum h_i/M(h_i) / \sum 1/M(h_i)$ can be taken to test the hypothesis that all H_i

genotypes have a common harvest index. The asymptotic distribution of Q when $(H_1 = \cdots = H_v)$ is χ^2 with v−1 degrees of freedom and the common harvest index is estimated by $\bar{H}$, the pooled estimate.

We take here the data on pod yield and total yield of 22 groundnut cultivars grown in three randomised blocks in 1984 at ICRISAT Center. We present the necessary computation on the analysis of harvest indices. In addition, a comparison of the proposed estimate with conventional estimate is elaborated with the help of a set of experimental data on groundnut.

RESULTS AND DISCUSSION

The analysis of variance for pod yield, total yield, estimates of error variances and correlations and χ^2 values are given in Table 1, and mean pod yield, total yield, and harvest indices (conventional and proposed), and ranking of genotypes based on the two types of harvest indices of the genotypes along with their biases and standard errors in Table 2. It can be seen that the biases are negligible in estimating harvest index from the data. The standard errors computed using approximation up to order (1/s) are reasonably close to those up to order $(1/s^2)$, at any rate up to three decimal places. The differences in the two approximations will decrease with increasing number of replications. It can be noted that in the present example, the percentage difference between the two estimates of harvest index $(\%d_i = 100(1- h_i\sqrt{r_i}))$ exceeds 22% for cultivar 19. The ranks of some genotypes are also affected.

Table 1. Mean squares for pod yield (X), total … of χ^2 for

Source	d.f.	MS X	MS Y
Blocks	2	7118	21557
Genotypes	21	4034	18404
	42	4735+	16031++
		MSP=7034	β = 0.81
	Q_1 = 28.867	Q_2 = 27.78	d.f. = 21

+, ++ estimates of σ^2_x and σ^2_y, respectively.
MS—mean squares; MSP—mean sum of product; d.f.—degrees of freedom; and Q_1, Q_2—values of Q (statistics for testing equality of harvest indices) (defined in the text).

While applying the test to the data of the experiment under discussion, we get the values Q_1 and Q_2 (Table 1) while using variances up to orders (1/s) and $(1/s^2)$, respectively. On comparing Q against the table value of χ^2 at 21 degrees of

Table 2. Mean pods yield and total yield (kg/ha), harvest indices (upper value h_i, lower value $\bar{r}_i$), per cent difference ($\%d_i$), ranks, biases and standard errors up to orders (1/s) and ($1/s^2$) for 22 genotypes

Genotype	Mean yield		Index	$\%d_i$	Rank	Bias × 10^4		SE × 10^4	
	pods	total				b_1	b_2	se_1	se_2
CGC 4063	279	751	0.372+	0.0	17+	−6+	−7+	319+	324+
			0.372++		17++	−21++	−27++	324++	325++
J 11 × Robut 331–1	236	555	0.425	1.8	7	−2	−3	422	433
			0.433		7	−8	−20	433	437
ICGS 24	279	789	0.354	−3.2	19	−7	−8	307	312
			0.343		20	−24	−30	312	313
ICGS 35	276	618	0.447	−0.2	3	−1	1	379	387
			0.446		3	4	−6	387	389
ICGS 11	331	684	0.484	2.3	1	5	5	346	352
			0.495		1	17	9	352	354
ICGS 24	235	719	0.328	2.1	21	−12	−12	345	352
			0.335		21	−38	−45	352	354
ICGS 21	251	646	0.388	1.5	15	−7	−7	367	375
			0.394		12	−22	−31	375	377
X 41-x-1-B×Goldin 1	274	698	0.393	−0.2	14	−5	−5	340	345
			0.392		14	−17	−25	345	347
Manfredi × X-14-4-B-19-B	253	604	0.419	1.2	9	−3	−3	389	397
			0.425		8	−10	−20	397	400
TMV 2	307	699	0.439	0.1	5	0	0	336	341
			0.439		5	0	−8	341	343
Faizpur 1-5-2	283	702	0.402	−2.6	12	−4	−4	336	341
			0.392		13	−13	−21	341	343
J 11	240	625	0.385	−0.4	16	−7	−8	381	389
			0.384		16	−25	−35	389	391
NCAC 17090	254	607	0.419	−0.2	10	−3	−3	387	395
			0.417		10	−10	−20	395	398
NCAC 17142	252	545	0.462	0.8	2	4	−5	431	443
			0.466		2	15	2	443	446
Gangapuri	347	793	0.438	0.2	6	0	0	296	299
			0.439		6	0	−6	299	300
EC 76446	239	607	0.394	−1.4	13	−7	−7	390	399
			0.389		15	−22	−32	399	401
EC 109271 (55-437)	290	793	0.366	−0.2	18	−6	−6	303	307
			0.365		18	−20	−26	307	308
EC 210 24	291	660	0.441	0.1	4	0	0	355	362
			0.441		4	1	−8	362	364
					17				
Manfredi 107	195	598	0.326	−22.5	22	−17	−17	415	426
			0.266		22	−57	−67	426	430
Krapovicas str. 16	216	524	0.413	0.8	11	−5	−5	449	462
			0.417		11	−18	−31	462	466
NCAC 16129	231	639	0.420	−0.2	8	−2	−2	351	357
			0.419		9	−8	−16	357	359
IL 21	219	627	0.349	0.7	20	−12	−13	388	397
			0.352		19	−41	−51	397	400
SE	±39.7	±73.1							

$\%d_i = 100\ (1-h_i/\bar{r}_i)$.

\+ Estimate of harvest index, rank of genotype based on harvest index, b_1 and b_2 [biases up to order (1/s) and ($1/s^2$)], se_1 and se_2 [standard errors up to order (1/s) and ($1/s^2$)] using proposed estimator h_i.

\++ Same as in +, but using conventional estimator $\bar{r}_i$.

freedom and 5% level of significance, the harvest indices of the genotypes do not show any significant departure from a common value

The data of some other trials were also analysed. There biases were also negligible and the standard errors from considering the terms up to orders (1/s) and ($1/s^2$) showed similar closeness as above (12).

The proposed method has provided more precise estimates of harvest indices compared to the conventional estimates. The difference in the two estimates, although small for the groundnut data, may be remarkable for other data sets. The computations in the proposed methods are straight forward, as we generally evaluate means of genotypes and their standard errors. However, one requires the estimation of residual correlation coefficient which is easily available from bivariate or multivariate analysis of variance procedures in standard statistical packages. Based on the above analysis, we recommend the estimation of harvest index as ratio of means instead of mean of ratios.

ACKNOWLEDGEMENTS

The author thanks Drs. K. N. Rai and J. H. Williams, ICRISAT, for encouraging comments, and Dr. R. C. Nageswara Rao, ICRISAT, for making data available for illustration in this paper.

REFERENCES

1. V. Arunachalam and A. Bandyopadhyay. 1979. Are "Multiple cross-Multiple Pollen Hybrids" an answer for productive population in *Brassica campestris* var. 'brown sarson'? I. Method for studying 'Mucromphs'. Theor. Apple. Genet., **54**: 203–207.

2. A. Bandyopadhyay and V. Arunachalam. 1982. Are "Multiple cross-Multiple Pollen Hybrids" an answer for productive population in *Brassica campestris* var. 'brown sarson'? III. Potential of component characters in population breeding. Theor. Appl. Genet., **61**: 53–58.

3. M. N. Murthy and N. S. Nanjamma. 1959. Almost unbiased ratio-estimates based on inter-penetrating sub-sample estimates. Sankhyā, **21**: 381–392.

4. W. Eisenhart. 1947. The assumptions underlying analysis of variance. Biometrics, **3**: 1–21.

5. W. G. Cochran. 1947. Some consequences when the assumptions for the analysis of variances are not satisfied. Biometrics, **3**: 22–38.

6. R. J. Baker. 1982. Effect of seeding rate on grain yield, straw yield and harvest index of eight spring wheat cultivars. Can. J. Pl. Sci., **62**: 285–291.

7. A. A. Rosielle and K. J. Frey. 1975. Estimates of selection parameters associated with harvest index in oat lines derived from a bulk population. Euphytica, **24**: 121–131.

8. I. D. Singh and N. C. Stoskoff. 1971. Harvest index in cereals. Agron. J., **63**: 224–226.

9. P. V. Sukhatme and B. V. Sukhatme. 1970. Sampling Theory of Surveys with Applications. Iowa University Press, Ames, Iowa, USA: 224–226.

10. M. G. Kendall and A. Stuart. 1979. The Advanced Theory of Statistics, Vol. 2. Charles Griffin. and Company Ltd., London: 748.

11. C. R. Rao. 1973. Linear Statistical Inference and Its Application. Wiley Eastern, New Delhi: 389.

12. M. Singh. 1985. Statistical Analysis of Ratios: 1. Harvest Indices. Statistics Unit Report No. 8/85 (unpublished). Statistics Unit, ICRISAT, Patancheru, A.P. India: 33.